Concentration of quadratic forms under a Bernstein moment assumption

Pierre C. Bellec *

ENSAE,
3 avenue Pierre Larousse,
92245 Malakoff Cedex, France.

Abstract: A concentration result for quadratic form of independent subgaussian random variables is derived. If the moments of the random variables satisfy a "Bernstein condition", then the variance term of the Hanson-Wright inequality can be improved.

1. Concentration of a quadratic form of subgaussian random variables

Throughout this note, $A \in \mathbb{R}^{n\times n}$ is a real matrix, and $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^T$ is a centered random vector with independent components. We are interested in the concentration behavior of the random variable

$$\boldsymbol{\xi}^T A \boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A \boldsymbol{\xi}],$$

Let $\sigma_i^2 = \mathbb{E}[\xi_i^2]$ for all $i = 1, ..., n$ and define $D_\sigma = \operatorname{diag}(\sigma_1, ..., \sigma_n)$. If the random variables $\xi_1, ..., \xi_n$ are Gaussian, we have the following concentration inequality.

Proposition 1 (Gaussian chaos of order 2)**.** *Let $\xi_1, ..., \xi_n$ be independent zero-mean normal random variables with for all $i = 1, ..., n$, $\mathbb{E}[\xi_i^2] = \sigma_i^2$. Let A be any $n \times n$ real matrix. Then for any $x > 0$,*

$$\mathbb{P}\left(\boldsymbol{\xi}^T A \boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A \boldsymbol{\xi}] > 2\left\|D_\sigma A D_\sigma\right\|_{\mathrm{F}} \sqrt{x} + 2|||D_\sigma A D_\sigma|||_2 x\right) \le \exp(-x).$$

A proof of this concentration result can be found in [3, Example 2.12]. We will refer to the term $2\left\|D_\sigma A D_\sigma\right\|_{\mathrm{F}} \sqrt{x}$ as the variance term, since if A is diagonal-free, the random variable $\boldsymbol{\xi}^T A \boldsymbol{\xi}$ is centered with variance

$$\left\|D_\sigma A D_\sigma\right\|_{\mathrm{F}}^2.$$

A similar concentration result is available for subgaussian random variables. It is known as the Hanson-Wright inequality and is given in Proposition 2 below. First versions of this inequality can be found in Hanson and Wright [5] and Wright [9], although with a weaker statement than Proposition 2 below since these results involve $|||\left(|a_{ij}|\right)|||_2$ instead of $|||A|||_2$. Recent proofs of this concentration inequality with $|||A|||_2$ instead of $|||\left(|a_{ij}|\right)|||_2$ can be found in Rudelson and Vershynin [6] or Barthe and Milman [2, Theorem A.5].

Proposition 2 (Hanson-Wright inequality [6])**.** *There exist an absolute constant $c > 0$ such that the following holds. Let $n \ge 1$ and $\xi_1, ..., \xi_n$ be independent zero-mean*

* This work was supported by GENES and by the grant Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

subgaussian random variables with $\max_{i=1,...,n} \|\xi_i\|_{\psi_2} \leq K$ *for some real number* $K > 0$. *Let* A *be any* $n \times n$ *real matrix. Then for all* $t > 0$,

$$\mathbb{P}\left(\boldsymbol{\xi}^T A\boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A\boldsymbol{\xi}] > t\right) \leq \exp\left(-c \min\left(\frac{t^2}{K^4 \|A\|_{\mathrm{F}}^2}, \frac{t}{K^2 |||A|||_2}\right)\right) \tag{1}$$

where $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^T$. *Furthermore, for any* $x > 0$, *with probability greater than* $1 - \exp(-x)$,

$$\boldsymbol{\xi}^T A\boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A\boldsymbol{\xi}] \leq cK^2 |||A|||_2 x + cK^2 \|A\|_{\mathrm{F}} \sqrt{x}.$$

For some random variables $\xi_1, ..., \xi_n$, the "variance term" $K^2 \|A\|_{\mathrm{F}} \sqrt{x}$ is far from the variance of the random variable $\boldsymbol{\xi}^T A\boldsymbol{\xi}$. The goal of the present paper is to show that under a mild assumption on the moments of $\xi_1, ..., \xi_n$, it is possible to substantially reduce the variance term. This assumption is the following.

Assumption 1 (Bernstein condition on $\xi_1^2, ..., \xi_n^2$)**.** *Let* $K > 0$ *and assume that* $\xi_1, ..., \xi_n$ *are independent and satisfy*

$$\forall p \geq 1, \qquad \mathbb{E}|\xi_i|^{2p} \leq \tfrac{1}{2}\, p!\, \sigma_i^2\, K^{2(p-1)}. \tag{2}$$

Example 1. Centered variables almost surely bounded by K and zero-mean Gaussian random variables with variance smaller than K^2 satisfy (2).

Example 2 (Log-concave random variables)**.** In [7], the authors consider a slightly stronger condition [7, Definition 1.1]. They consider random variables Z satisfying for any integer $p \geq 1$ and some constant K:

$$\mathbb{E}[|Z|^p] \leq\ p\, K\, \mathbb{E}[|Z|^{p-1}], \tag{3}$$

and they showed in [7, Section 7] that any distribution that is log-concave satisfies (3). Thus, if X^2 is log-concave then our assumption (2) holds. See [1, Section 6] for a comprehensive list of the common log-concave distributions.

The next theorem provides a concentration inequality for quadratic forms of independent random variables satisfying the moment assumption (2). It is sharper than the Hanson-Wright inequality given in Proposition 2.

Theorem 3. *Assume that the random variable* $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^T$ *satisfies Assumption 1 for some* $K > 0$. *Let* A *be any* $n \times n$ *real matrix. Then for all* $t > 0$,

$$\mathbb{P}\left(\boldsymbol{\xi}^T A\boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A\boldsymbol{\xi}] > t\right) \leq \exp\left(-\min\left(\frac{t^2}{192K^2 \|AD_\sigma\|_{\mathrm{F}}^2}, \frac{t}{256K^2 |||A|||_2}\right)\right), \tag{4}$$

where $D_\sigma = \operatorname{diag}(\sigma_1, ..., \sigma_n)$. *Furthermore, for any* $x > 0$, *with probability greater than* $1 - \exp(-x)$,

$$\boldsymbol{\xi}^T A\boldsymbol{\xi} - \mathbb{E}[\boldsymbol{\xi}^T A\boldsymbol{\xi}] \leq 256K^2 |||A|||_2 x + 8\sqrt{3}K \|AD_\sigma\|_{\mathrm{F}} \sqrt{x}. \tag{5}$$

The proof of this result relies on the decoupling inequality for quadratic forms [8] [4, Theorem 8.11].

If t is small, the right hand side of (4) becomes

$$\exp\left(-\frac{t^2}{192K^2 \|AD_\sigma\|_{\mathrm{F}}^2}\right),$$

whereas the right hand side of the Hanson-Wright inequality (1) becomes

$$\exp\left(-c\frac{t^2}{K^4\,\|A\|_{\mathrm{F}}^2}\right),$$

for some absolute constant $c > 0$. The element of the diagonal matrix D_σ are bounded from above by K, so Theorem 3 gives a sharper bound than the Hanson-Wright inequality in this regime.

2. Proof of Theorem 3

The goal of this section is to prove Theorem 3. We start with preliminary calculations that will be useful in the proof. Let A be any $n \times n$ real matrix. Let $\lambda > 0$ satisfy

$$128|||A|||_2 K^2\lambda \leq 1, \tag{6}$$

and define

$$\eta = 32K^2\lambda^2. \tag{7}$$

The inequality (6) can be rewritten in terms of η:

$$512K^2|||A|||_2^2\eta \leq 1. \tag{8}$$

Let A_0 be the matrix A with the diagonal entries set to 0. Then, using the triangle inequality with $A_0 = A - \mathrm{diag}(a_{11}, ..., a_{nn})$ and $|a_{ii}| \leq |||A|||_2$ for all $i = 1, ..., n$, we obtain

$$|||A_0|||_2 \leq 2|||A|||_2. \tag{9}$$

Let $B = A_0^T A_0 = (b_{ij})_{i,j=1,...,n}$ and let B_0 be the matrix B with the diagonal entries set to 0. Then

$$\forall i = 1, ..., n, \qquad 0 \leq b_{ii} = \sum_{j\neq i} a_{ji}^2 \leq |||A|||_2^2. \tag{10}$$

By using the decomposition $B_0 = B - \mathrm{diag}(b_{11}, ..., b_{nn})$ and the inequality $\|v + v'\|_2^2 \leq 2\|v\|_2^2 + 2\|v'\|_2^2$, (10) and (9), we have:

$$\begin{aligned}
\|B_0\boldsymbol{\xi}\|_2^2 &\leq 2\|B\boldsymbol{\xi}\|_2^2 + 2\sum_{i=1}^n b_{ii}^2\xi_i^2, \\
&\leq 2|||A_0|||_2^2\|A_0\boldsymbol{\xi}\|_2^2 + 2|||A|||_2^2\sum_{i=1}^n b_{ii}\xi_i^2, \\
&\leq 8|||A|||_2^2\|A_0\boldsymbol{\xi}\|_2^2 + 2|||A|||_2^2\sum_{i=1}^n b_{ii}\xi_i^2.
\end{aligned}$$

Combining the previous display with (8), we obtain for any $K > 0$:

$$\begin{aligned}
16K^2\eta^2\|B_0\boldsymbol{\xi}\|_2^2 &\leq (512K^2|||A|||_2^2\eta)\left(\frac{\eta}{4}\|A_0\boldsymbol{\xi}\|_2^2 + \frac{\eta}{16}\sum_{i=1}^n b_{ii}\xi_i^2\right), \\
&\leq \frac{\eta}{4}\|A_0\boldsymbol{\xi}\|_2^2 + \frac{\eta}{16}\sum_{i=1}^n b_{ii}\xi_i^2.
\end{aligned} \tag{11}$$

Proof of Theorem 3. Throughout the proof, let $\lambda > 0$ satisfy (6). The value of λ will be specified later.

First we treat the diagonal terms by bounding the moment generating function of

$$S_{\text{diag}} \coloneqq \sum_{i=1}^n a_{ii}\xi_i^2 - \sum_{i=1}^n a_{ii}\sigma_i^2.$$

Using the independence of $\xi_1, ..., \xi_n$ and (17) with $s = a_{ii}\lambda$ with each $i = 1, ..., n$:

$$\mathbb{E}\exp(\lambda S_{\text{diag}}) \leq \exp\left(\lambda^2 \sum_{i=1}^n a_{ii}^2\sigma_i^2 K^2\right), \tag{12}$$

provided that for all $i = 1, ..., n$, $2|a_{ii}|\lambda K^2 \leq 1$ which is satisfied as (6) holds and $|a_{ii}| \leq |||A|||_2$.

Now we bound the moment generating function of the off-diagonal terms. Let

$$S_{\text{off}-\text{diag}} \coloneqq \sum_{i,j=1,...,n: i\neq j} a_{ij}\xi_i\xi_j.$$

Let the random vector $\boldsymbol{\xi}' = (\xi_1', ..., \xi_n')^T$ be independent of $\boldsymbol{\xi}$ with the same distribution as $\boldsymbol{\xi}$. We apply the decoupling inequality [8] (see also [4, Theorem 8.11]) to the convex function $s \to \exp(\lambda s)$:

$$\mathbb{E}\exp(\lambda S_{\text{off}-\text{diag}}) \leq \mathbb{E}\exp\left(4\lambda \sum_{i,j=1,...,n: i\neq j} a_{ij}\xi_i'\xi_j\right).$$

Conditionally on $\xi_1, ..., \xi_n$, for each $i = 1, ..., n$, we use the independence of $\xi_1', ..., \xi_n'$ and (16) applied to ξ_i' with $s = 4\sum_{j=1,...,n: i\neq j} a_{ij}\xi_j$:

$$\begin{aligned}\mathbb{E}\exp\left(4\lambda\sum_{i\neq j} a_{ij}\xi_i'\xi_j\right) &\leq \mathbb{E}\exp\left(16K^2\lambda^2 \sum_{i=1,...,n}\left(\sum_{j=1,...,n: i\neq j} a_{ij}\xi_j\right)^2\right), \\ &= \mathbb{E}\exp\left(16K^2\lambda^2\|A_0\boldsymbol{\xi}\|_2^2\right) = \mathbb{E}\exp\left(\frac{\eta}{2}\|A_0\boldsymbol{\xi}\|_2^2\right),\end{aligned}$$

where η is defined in (7) and A_0 is the matrix A with the diagonal entries set to 0. Let $B = A_0^T A_0 = (b_{ij})_{i,j=1,...,n}$. Then $\|A_0\boldsymbol{\xi}\|_2^2 = \sum_{i=1}^n b_{ii}\xi_i^2 + \sum_{i\neq j} b_{ij}\xi_i\xi_j$.

We use the Cauchy-Schwarz inequality to separate the diagonal terms from the off-diagonal ones:

$$\left(\mathbb{E}\exp(\frac{\eta}{2}\|A_0\boldsymbol{\xi}\|_2^2)\right)^2 \leq \mathbb{E}\exp\left(\eta\sum_{i=1}^n b_{ii}\xi_i^2\right)\mathbb{E}\exp\left(\eta\sum_{i\neq j} b_{ij}\xi_i\xi_j\right). \tag{13}$$

For the off-diagonal terms of (13), using the decoupling inequality [8] (see also [4, Theorem 8.11]) we have:

$$\mathbb{E}\exp\left(\eta\sum_{i\neq j} b_{ij}\xi_i\xi_j\right) \leq \mathbb{E}\exp\left(4\eta\sum_{i\neq j} b_{ij}\xi_i'\xi_j\right).$$

Again, conditionally on $\xi_1,...,\xi_n$, for each $j=1,...,n$, we use (16) applied to ξ_i' and the independence of $\xi_1',...,\xi_n'$:

$$\begin{aligned}\mathbb{E}\exp\left(4\eta\sum_{i\neq j} b_{ij}\xi_i'\xi_j\right) &\leq \mathbb{E}\exp\left(16K^2\eta^2\sum_{i=1}^n\left(\sum_{j=1,\dots,n:\ i\neq j} b_{ij}\xi_j\right)^2\right),\\ &= \mathbb{E}\exp\left(16K^2\eta^2\|B_0\boldsymbol{\xi}\|_2^2\right),\\ &\leq \mathbb{E}\exp\left(\frac{\eta}{4}\|A_0\boldsymbol{\xi}\|_2^2 + \frac{\eta}{16}\sum_{i=1}^n b_{ii}\xi_i^2\right),\end{aligned}$$

where we used the preliminary calculation (11) for the last display. Finally, the Cauchy-Schwarz inequality yields

$$\mathbb{E}\exp\left(4\eta\sum_{i\neq j} b_{ij}\xi_i\xi_j'\right) \leq \sqrt{\mathbb{E}\exp\left(\frac{\eta}{2}\|A_0\boldsymbol{\xi}\|_2^2\right)}\sqrt{\mathbb{E}\exp\left(\frac{\eta}{8}\sum_{i=1}^n b_{ii}\xi_i^2\right)}.$$

We plug this upper bound back into (13). After rearranging, we find

$$\left(\mathbb{E}\exp(\frac{\eta}{2}\|A_0\boldsymbol{\xi}\|_2^2)\right)^{3/2} \leq \mathbb{E}\exp\left(\eta\sum_{i=1}^n b_{ii}\xi_i^2\right)\sqrt{\mathbb{E}\exp\left(\frac{\eta}{8}\sum_{i=1}^n b_{ii}\xi_i^2\right)}.$$

As $b_{ii}\geq 0$, this implies:

$$\mathbb{E}\exp(\frac{\eta}{2}\|A_0\boldsymbol{\xi}\|_2^2) \leq \mathbb{E}\exp\left(\eta\sum_{i=1}^n b_{ii}\xi_i^2\right).$$

For each $i=1,...,n$, we apply (18) to the variable ξ_i with $s=b_{ii}\eta\geq 0$. Using the independence of $\xi_1^2,...,\xi_n^2$, we obtain:

$$\begin{aligned}\mathbb{E}\exp\left(\eta\sum_{i=1}^n b_{ii}\xi_i^2\right) &= \prod_{i=1}^n \mathbb{E}\exp(\eta b_{ii}\xi_i^2),\\ &\leq \exp\left(\frac{3}{2}\eta\sum_{i=1}^n b_{ii}\sigma_i^2\right) = \exp\left(\frac{3}{2}\eta\,\|A_0D_\sigma\|_{\mathrm{F}}^2\right).\end{aligned}$$

provided that for all $i=1,...,n$, $2K^2b_{ii}\eta\leq 1$ which is satisfied thanks to (6) and (10).

We remove η from the above displays using its definition (7):

$$\mathbb{E}\exp(\lambda S_{\mathrm{off-diag}}) \leq \exp\left(48\lambda^2K^2\,\|A_0D_\sigma\|_{\mathrm{F}}^2\right), \tag{14}$$

where A_0 is the matrix A with the diagonal entries set to 0.

Now we combine the bound on the moment generating function of S_{diag} and $S_{\mathrm{off-diag}}$, given respectively in (12) and (14). Using the Chernoff bound and the Cauchy-Schwarz inequality: we have that for all λ satisfying (6),

$$\begin{aligned}\mathbb{P}\left(S_{\rm diag}+S_{\rm off-diag}>t\right) &\le \exp(-\lambda t)\mathbb{E}[\exp(\lambda S_{\rm diag})\exp(\lambda S_{\rm off-diag})],\\ &\le \exp\left(-\lambda t\right)\sqrt{\mathbb{E}[\exp(2\lambda S_{\rm diag})]}\sqrt{\mathbb{E}[\exp(2\lambda S_{\rm off-diag})]},\\ &\le \exp\left(-\lambda t+\lambda^2K^2\left(\sum_{i=1}^n \sigma_i^2 a_{ii}^2 + 48\left\|A_0D_\sigma\right\|_{\rm F}^2\right)\right),\\ &\le \exp\left(-\lambda t + 48\lambda^2K^2\left\|AD_\sigma\right\|_{\rm F}^2\right), \qquad (15)\end{aligned}$$

where for the last display we used the equality

$$\left\|AD_\sigma\right\|_{\rm F}^2 = \sum_{i,j=1,\dots,n} a_{ij}^2\sigma_i^2 = \left\|A_0D_\sigma\right\|_{\rm F} + \sum_{i=1}^n a_{ii}^2\sigma_i^2.$$

It now remains to choose the parameter λ. The unconstrained minimum of (15) is attained at $\bar\lambda = t/(96K^2\left\|AD_\sigma\right\|_{\rm F}^2)$. If $\bar\lambda$ satisfies the constraint (6), then

$$\mathbb{P}\left(S_{\rm diag}+S_{\rm off-diag}>t\right) \le \exp\left(\frac{-t^2}{192K^2\left\|AD_\sigma\right\|_{\rm F}^2}\right).$$

On the other hand, if $\bar\lambda$ does not satisfy (6), then the constraint (6) is binding and the minimum of (15) is attained at $\lambda_{\rm b} = 1/(128|||A|||_2K^2) < \bar\lambda$. In this case,

$$-t\lambda_{\rm b}+\lambda_{\rm b}^2 48K^2\left\|AD_\sigma\right\|_{\rm F}^2 \le -t\lambda_{\rm b}+\lambda_{\rm b}\bar\lambda 48K^2\left\|AD_\sigma\right\|_{\rm F}^2 = -t\lambda_{\rm b}+\frac{t}{2}\lambda_{\rm b} = -\frac{t}{256K^2|||A|||_2}.$$

Combining the two regimes, we obtain

$$\mathbb{P}\left(S_{\rm diag}+S_{\rm off-diag}>t\right) \le \exp\left(-\min\left(\frac{t^2}{192K^2\left\|AD_\sigma\right\|_{\rm F}^2}, \frac{t}{256K^2|||A|||_2}\right)\right).$$

The proof of (4) is complete.

Now we prove (5). The function

$$t \to x(t) = \min\left(\frac{t^2}{192K^2\left\|AD_\sigma\right\|_{\rm F}^2}, \frac{t}{256K^2|||A|||_2}\right)$$

is increasing and bijective from the set of positive real numbers to itself. Furthermore, for all $t>0$,

$$t \le 8\sqrt{3}K\left\|AD_\sigma\right\|_{\rm F}\sqrt{x(t)} + 256K^2|||A|||_2 x(t),$$

so the variable change $x = x(t)$ completes the proof of (5). □

3. Technical lemmas: bounds on moment generating functions

The condition (2) leads to the following bounds on the moment generating functions of X and X^2, which are crucial to prove Theorem 3.

Proposition 4. *Let $K > 0$ and let ξ_i be a random variable satisfying (2) with $\sigma_i^2 = \mathbb{E}[\xi_i^2]$. Then for all $s \in \mathbf{R}$:*

$$\mathbb{E}\exp(s\xi_i) \leq \exp(s^2 K^2). \tag{16}$$

Furthermore, if $0 \leq 2sK^2 \leq 1$, then

$$\mathbb{E}\exp(s\xi_i^2 - s\sigma_i^2) \leq \exp(s^2\sigma_i^2 K^2), \tag{17}$$

$$\mathbb{E}\exp(s\xi_i^2) \leq \exp\left(\frac{3}{2}s\sigma_i^2\right). \tag{18}$$

Inequality (16) shows that a random variable X satisfying the moment assumption (2) is subgaussian and its ψ_2 norm is bounded by K up to a multiplicative absolute constant. The proof of Proposition 4 is based on Taylor expansions and some algebra.

Proof of Proposition 4. To simplify the notation, let $X = \xi_i$ and $\sigma = \sigma_i$. We first prove (17). We apply the assumption on the even moments of X:

$$\begin{aligned}\mathbb{E}\exp(sX^2) &= 1 + s\sigma^2 + \sum_{p\geq 2} \frac{s^p \mathbb{E}X^{2p}}{p!}, \\ &\leq 1 + s\sigma^2 + \frac{\sigma^2 s}{2}\sum_{k=1}^{\infty}(sK^2)^k = 1 + s\sigma^2 + \frac{\sigma^2 K^2 s^2}{2(1-sK^2)},\end{aligned}$$

and using the inequality $0 < 2sK^2 \leq 1$, we obtain:

$$\mathbb{E}\exp(sX^2) \leq 1 + s\sigma^2 + \sigma^2 s^2 K^2 \leq \exp(s\sigma^2 + s^2\sigma^2K^2),$$

which completes the proof of (17). Inequality (18) is a direct consequence of (17) after applying again the inequality $2sK^2 \leq 1$.

We now prove (16). Using the Cauchy-Schwarz inequality and the assumption on the moments for $p = 2$, we get $\sigma^4 \leq \mathbb{E}[\xi^4] \leq \sigma^2 K^2$, so $\sigma \leq K$. Let $p \geq 1$. For the even terms of the expansion of $\mathbb{E}\exp(sX)$, we get:

$$\frac{s^{2p}\mathbb{E}X^{2p}}{(2p)!} \leq \tfrac{1}{2}(sK)^{2p}\frac{p!}{(2p)!} \leq \tfrac{1}{2}\frac{(sK)^{2p}}{p!},$$

where for the last inequality we used $(p!)^2 \leq (2p)!$. For the odd terms, by using the Jensen inequality for $p \geq 1$:

$$\begin{aligned}\frac{s^{2p+1}\mathbb{E}X^{2p+1}}{(2p+1)!} \leq \frac{s^{2p+1}(\mathbb{E}X^{2p+2})^{\frac{2p+1}{2p+2}}}{(2p+1)!} &\leq |sK|^{2p+1}\frac{\left(\frac{(p+1)!}{2}\right)^{\frac{2p+1}{2p+2}}}{(2p+1)!}, \\ &\leq \tfrac{1}{2}|sK|^{2p+1}\frac{(p+1)!}{(2p+1)!}.\end{aligned}$$

If $|sK| > 1$, we use the inequality $(p+1)!^2 \leq (2p+1)!$ to obtain

$$\frac{s^{2p+1}\mathbb{E}X^{2p+1}}{(2p+1)!} \leq \frac{|sK|^{2(p+1)}}{2((p+1)!)},$$

and by combining the inequality for the even and the odd terms:

$$
\begin{aligned}
\mathbb{E}\exp(sX) &= 1 + \sum_{p\geq 1} \frac{s^{2p}\mathbb{E}X^{2p}}{(2p)!} + \frac{s^{2p+1}\mathbb{E}X^{2p+1}}{(2p+1)!}, \\
&\leq 1 + \tfrac{1}{2}\sum_{p\geq 1} \frac{(sK)^{2p}}{p!} + \frac{|sK|^{2(p+1)}}{(p+1)!}, \\
&\leq 1 + \sum_{p\geq 1} \frac{(sK)^{2p}}{p!} = \exp(s^2K^2).
\end{aligned}
$$

If $|sK| \leq 1$, we use the inequality $(p+1)!p! \leq (2p+1)!$ to obtain

$$
\frac{s^{2p+1}\mathbb{E}X^{2p+1}}{(2p+1)!} \leq \frac{(sK)^{2p}}{2(p!)},
$$

and by combining the inequality for the even and the odd terms:

$$
\begin{aligned}
\mathbb{E}\exp(sX) &= 1 + \sum_{p\geq 1} \frac{s^{2p}\mathbb{E}X^{2p}}{(2p)!} + \frac{s^{2p+1}\mathbb{E}X^{2p+1}}{(2p+1)!}, \\
&\leq 1 + \tfrac{1}{2}\sum_{p\geq 1} \frac{(sK)^{2p}}{p!} + \frac{(sK)^{2p}}{p!} = 1 + \sum_{p\geq 1} \frac{(sK)^{2p}}{p!} = \exp(s^2K^2).
\end{aligned}
$$

□

References

[1] Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.

[2] Franck Barthe and Emanuel Milman. Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Comm. Math. Phys.*, 323(2):575–625, 2013. . URL http://dx.doi.org/10.1007/s00220-013-1782-2.

[3] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[4] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[5] David Lee Hanson and Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[6] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 1–9, 2013. URL http://ecp.ejpecp.org/article/view/2865.

[7] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012.

[8] Roman Vershynin. A simple decoupling inequality in probability theory. 2011. URL http://www-personal.umich.edu/~romanv/papers/decoupling-simple.pdf.

[9] Farrol Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1(6):1068–1070, 1973.